As in Scenario A, it may be possible to estimate p directly from remediation control survey data. To do this, the data should first be shifted so that the median is at the $DCGL_W$ before the proportion that falls above the LBGR is calculated.

p may also be estimated from the odds that a random measurement will be greater than the LBGR versus that it is below the LBGR *when the median is near the $DCGL_W$*. If these odds are $r_1:r_2$, then $p = r_1/(r_1+r_2)$. For example, if the odds that a random measurement is greater than the LBGR when the median is near the $DCGL_W$, are 3:1, then $p = 3/(1+3) = 3/4 = 0.75$. Notice that since we are assuming that the true median is near the $DCGL_W$ the odds must be greater than 1:1. Once a survey unit has been remediated, it may be somewhat unnatural to try to estimate the odds this way.

9.4 Sample Size Calculation for the WRS Test Under Scenario A

For the WRS test, the *total* number of required samples from the reference area and survey unit combined is estimated from (Noether, 1987):

$$N = \frac{(Z_{1-\alpha} + Z_{1-\beta})^2}{12\,c\,(1-c)(P_r - 0.5)^2} \tag{9-6}$$

where:
α = specified Type I error rate
β = specified Type II error rate
$Z_{1-\alpha}$ = $100(1-\alpha)$ percentile of the standard normal distribution function
$Z_{1-\beta}$ = $100(1-\beta)$ percentile of the standard normal distribution function
c = proportion of measurements taken in the survey unit.
P_r = estimated probability that a random measurement from the survey unit exceeds a random measurement from the reference area by less than the $DCGL_W$ when the survey unit median is at the LBGR above background. $P_r \neq 0.5$

The numerator of equation 9-6 is the same as that in equation 9-1. Therefore, the same methods are used to calculate it as were discussed in Section 9.2. Table 9.2 gives commonly used values of α and β, together with the corresponding values of $(Z_{1-\alpha} + Z_{1-\beta})^2$. For planning purposes, c is set equal to 0.5, so that Equation 9-6 becomes

$$N = \frac{(Z_{1-\alpha} + Z_{1-\beta})^2}{3\,(P_r - 0.5)^2} \tag{9-6'}$$

The denominator of Equation 9-6 differs from that in Equation 9-1 in three important ways. First, the parameter P_r replaces the parameter p. It is a different probability, but because it is still a probability, $0 \leq P_r \leq 1$. The factor $(P_r - 0.5)^2$ cannot be larger than 0.25. Second, the constant

factor in the denominator is 3 rather than 4. This means that the sample size multiplier $1/[3(P_r - 0.5)^2]$ cannot be smaller than 4/3. Third, Equation 9.6 yields the total number of samples required in both the survey unit and a reference area. *N/2* samples will be taken in each.

The definition of the parameter P_r states that it is the estimated probability that a random measurement from the survey unit exceeds a random measurement from the reference area by less than the $DCGL_W$ when the survey unit median is above the reference area median by an amount equal to the concentration value at the LBGR. This is illustrated in Figure 9.5.
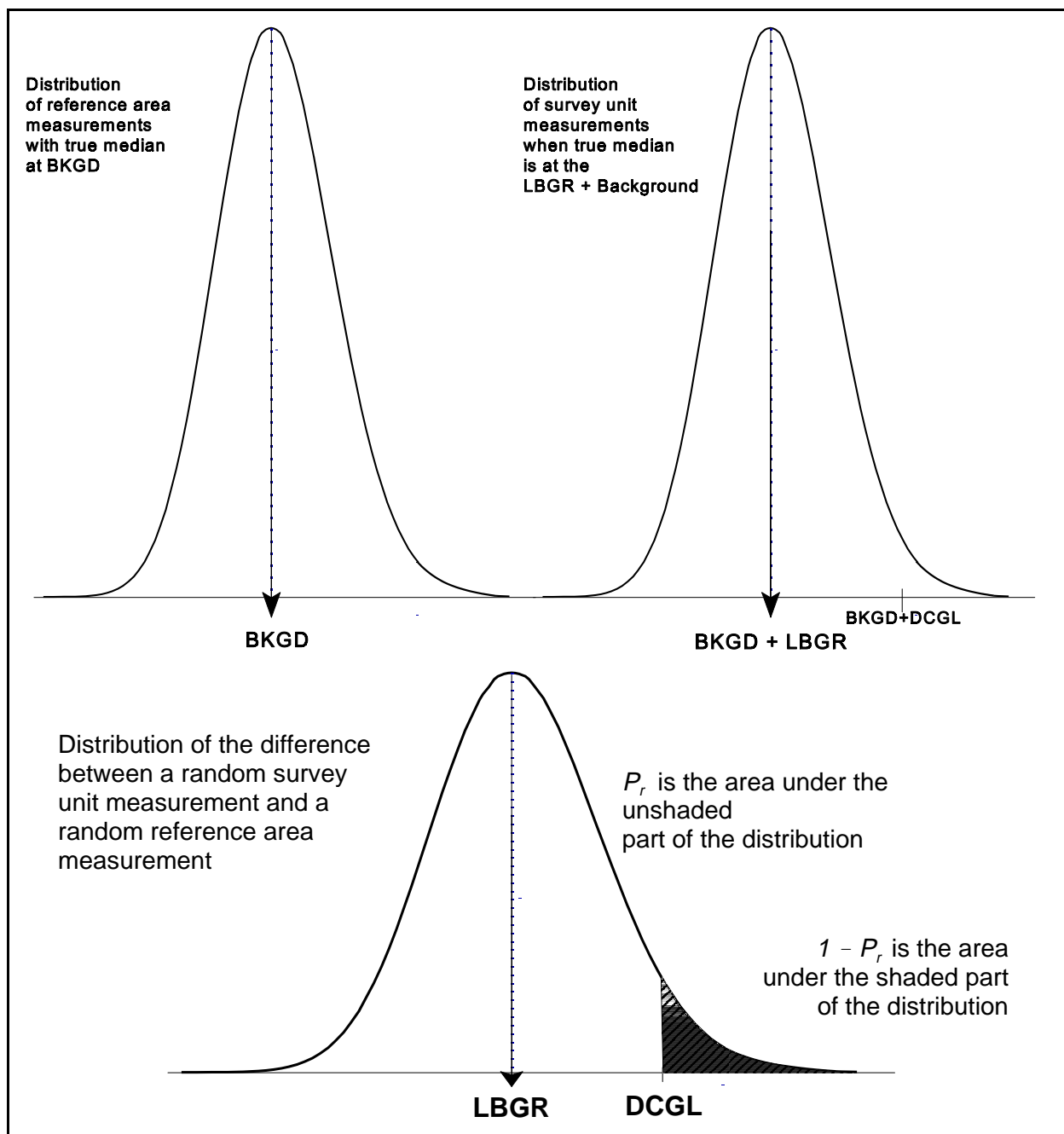

Figure 9.5 The Parameter P_r for the WRS Test Under Scenario A

As was done for the Sign test, the normal distribution may be used to facilitate the conversion of the values of Δ/σ to values of P_r in order to calculate the required sample sizes. The normal distribution is not used to actually conduct the test. Values of P_r are computed for a normal distribution from the following equation:

$$P_r = \text{Probability}(U = X - Y < DCGL)$$

$$= \int_{-\infty}^{DCGL} \left[\int_{-\infty}^{\infty} f_X(u + y) f_Y(y) dy \right] du$$

$$= \int_{-\infty}^{DCGL} \left[\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-(u+y-LBGR-BKGD)^2/2\sigma^2} \frac{1}{\sqrt{2\pi}\sigma} e^{-(y-BKGD)^2/2\sigma^2} dy \right] du$$

$$= \frac{1}{\sqrt{2\pi}\sqrt{2}\sigma} \int_{-\infty}^{DCGL} e^{-(u-LBGR)^2/4\sigma^2} du$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{DCGL_W - LBGR}{\sqrt{2}\sigma}} e^{-x^2/2} dx$$

$$= \Phi\left(\frac{\Delta}{\sqrt{2}\sigma} \right) \tag{9-7}$$

Values of P_r as a function of Δ/σ are listed in Table 9.4. The probability P_r, and the sample size multiplier $1/[3(P_r - 0.5)^2]$ are shown as a function of Δ/σ in Figure 9.6.

Table 9.4 Values of P_r for Use in Computing Sample Size for the WRS Test

| Δ/σ	P_r	Δ/σ	P_r	Δ/σ	P_r	Δ/σ	P_r
0.1	0.528186	1.1	0.781662	2.1	0.931218	3.1	0.985811
0.2	0.556231	1.2	0.801928	2.2	0.940103	3.2	0.988174
0.3	0.583998	1.3	0.821015	2.3	0.948062	3.3	0.990188
0.4	0.611351	1.4	0.838901	2.4	0.955157	3.4	0.991895
0.5	0.638163	1.5	0.855578	2.5	0.961450	3.5	0.993336
0.6	0.664313	1.6	0.871050	2.6	0.967004	4.0	0.997661
0.7	0.689691	1.7	0.885334	2.7	0.971881	5.0	0.999796
0.8	0.714196	1.8	0.898454	2.8	0.976143	6.0	0.999989
0.9	0.737741	1.9	0.910445	2.9	0.979848		
1.0	0.760250	2.0	0.921350	3.0	0.983053		

NUREG-1505
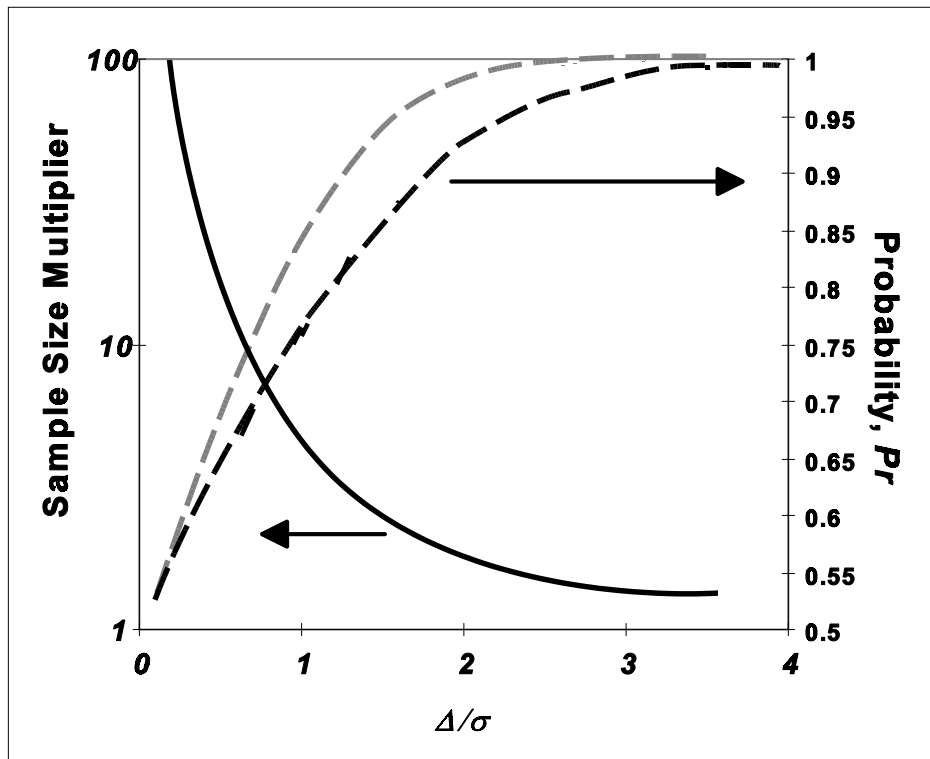
Figure 9.6 Dependence of Sample Size on Δ/σ for WRS Test
(Values of *p* used for the Sign Test are shown in gray for comparison)

Notice that the only difference between Equation 9-7, and equation 9-4 for computing *p* for the Sign test is that the standard deviation σ is replaced by $\sqrt{2}$ times σ. This is because the variance of the difference of two independent measurements is the sum of the variances of the individual measurements. If the variances of the individual measurements are about the same, i.e., σ^2, the variance of their difference is $2\sigma^2$. Thus, for a given value of Δ/σ, P_r will always be less than *p*. This also causes the total sample size required for the WRS test to be greater than that for the Sign test. Values of *p* are shown for comparison to P_r by the gray dashed line in Figure 9.6.

The combined effect of all the differences between Equation 9-7 and 9-4 is summarized in Figure 9.7.

Values of P_r for distributions other than normal can be calculated from the following equation:

$$P_r = \text{Probability}(U = X - Y < DCGL) = \int_{-\infty}^{DCGL}\left[\int_{-\infty}^{\infty} f_X(u + y)f_Y(y)dy\right]du \qquad (9\text{-}8)$$

where *Y* is a random measurement from the reference area with density f_Y and *X* is a random measurement from the survey unit with density f_X. However, in PNL-8989 (1993), Hardin and Gilbert have found that using the values of P_r from Equation 9-6 yielded good results when the

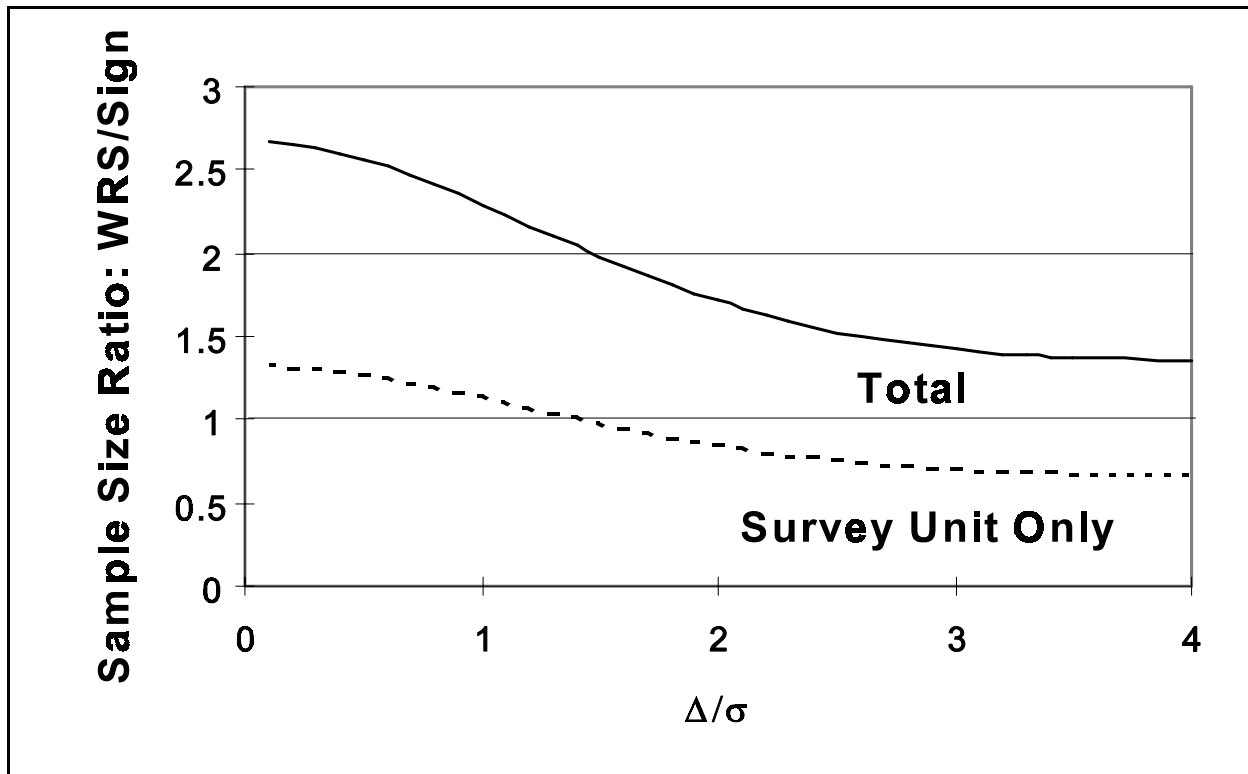distributions being tested were positively skewed, such as the log-normal.

Figure 9.7 Comparison of Sample Sizes Required for the WRS Test and the Sign Test

In some situations, it may be possible to estimate P_r directly from remediation control survey data. It is an estimate of the proportion of time that a random survey measurement will exceed a random reference area measurement by less than the $DCGL_W$.

P_r may also be estimated by the odds that a random survey measurement will exceed a random reference area measurement by less than the $DCGL_W$ versus that a random survey measurement will exceed a random reference area measurement by more than the $DCGL_W$. If these odds are $r_1 : r_2$, then $P_r = r_1 / (r_1 + r_2)$.

Whatever method is used to estimate P_r, it is important not to overestimate it, since that will result in a sample size inadequate to achieve the desired power of the test. The dependence of the sample size multiplier, $1/[3(P_r - 0.5)^2]$, on P_r is shown in Figure 9.8.

As an illustration, consider the example given in Chapter 6.1. For that example, the $DCGL_W = 160$, the $LBGR = 142$, $\alpha = \beta = 0.05$, and $\sigma = 6$. From Table 9.2,$(Z_{1-\alpha} + Z_{1-\beta})^2 = 11$ when $\alpha = \beta = 0.05$. The width of the gray region, $\Delta = DCGL_W - LBGR = 160 - 142 = 18$, so $\Delta/\sigma = 18/6 = 3$.

From Table 9.4, the value of P_r using the normal approximation is 0.983053. Thus the factor $1/[3(p - 0.5)^2] = 1/[3(0.983053 - 0.5)^2]$

$$= 1/[3(0.483053)^2]$$
$$= 1/[3(0.233340)]$$
$$= 1/0.700021$$
$$\approx 1.43.$$

So, the minimum sample size of 11 is increased by a factor of 1.43 to 15.7. This would normally be rounded up to 16, or 8 samples each in the reference area and the survey unit. However, because Equation 9-6 is an approximation, it is prudent to increase this number moderately. An increase of 20% is recommended. This increases the number of samples to $1.2(15.7) = 18.9$, which is rounded up to the next even integer, 20. Thus, 10 samples each in the reference area and the survey unit are required. This is the number that appears in Table 3.3.
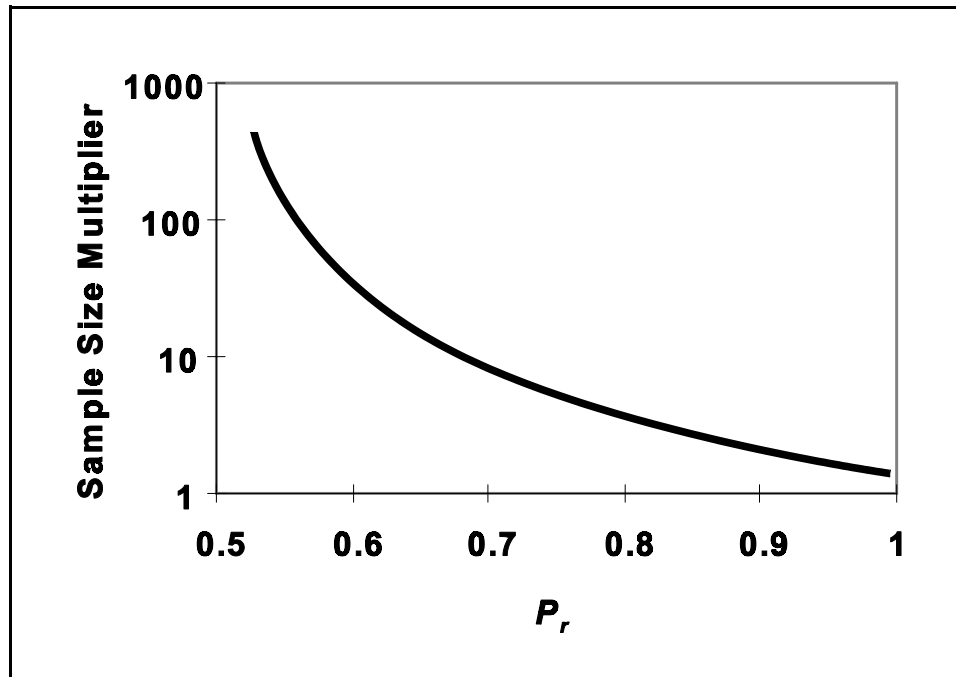

Figure 9.8 Dependence of Sample Size on P_r

9.5 Sample Size Calculation for the WRS Test Under Scenario B

Under Scenario B, Equation 9-6 is also used to estimate the required sample size. The roles of α and β are reversed, but this has no effect on the numerator of Equation 9-6, so Table 9.2 may still be used. However, since under Scenario B, both the WRS test and the Quantile test are used in tandem, the value of α decided on during the DQO process is halved for each test. Thus, the Table 9.2 value for $\alpha_W = \alpha/2$ and β is used.

The form of the denominator also remains the same, and Figure 9.8 still represents the dependence of the sample size multiplier on P_r. However, the definition of the parameter P_r is different.   The definition of the parameter P_r under Scenario B is the estimated probability that the difference between a random measurement from the survey unit and a random measurement from the reference area will be greater than the LBGR when the survey unit median is actually at the $DCGL_W$ above the background median. This is illustrated in Figure 9.9.
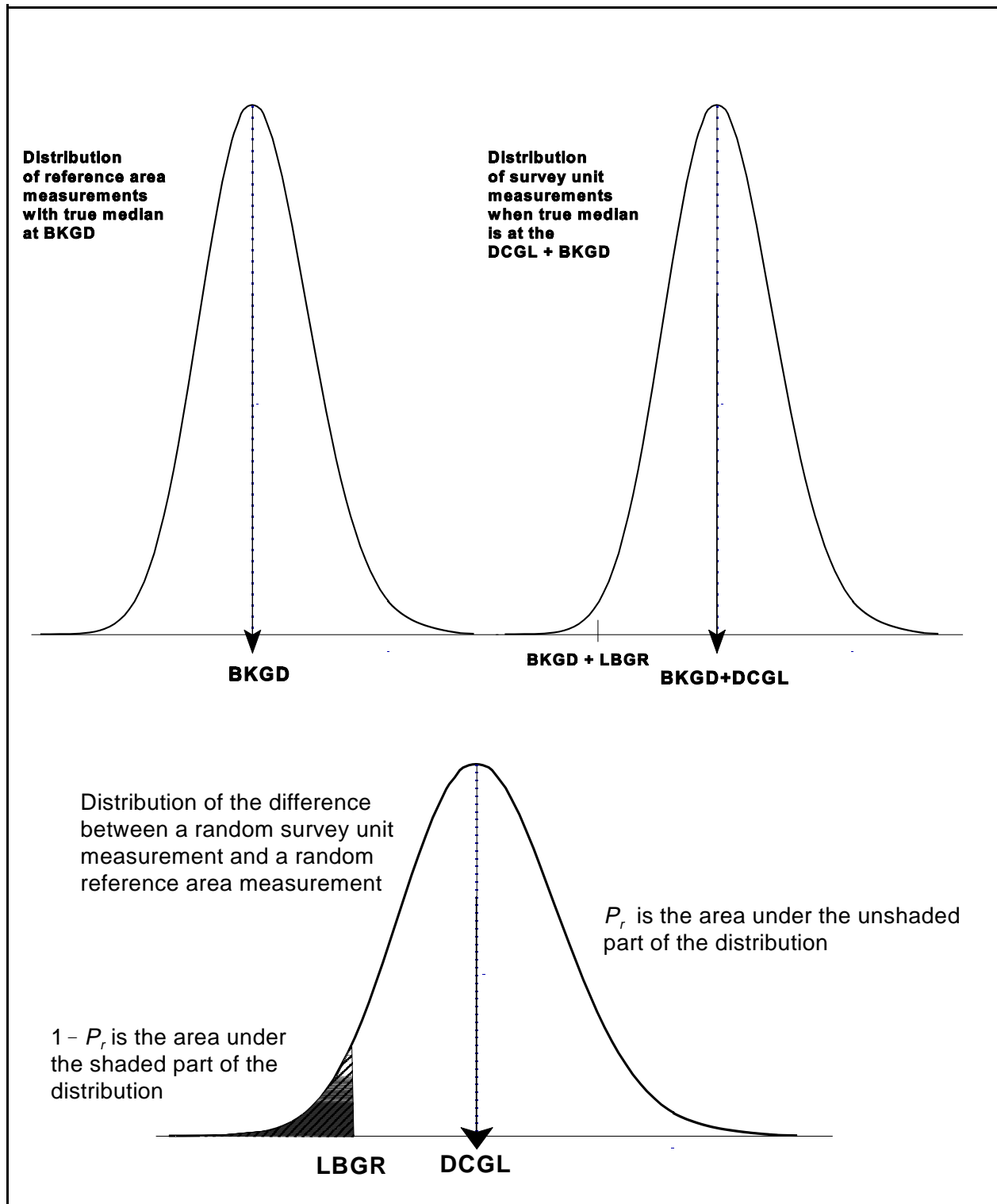
Figure 9.9 The Parameter P_r for the WRS Test Under Scenario B

The value of $1 - P_r$ expresses the likelihood that differences less than the LBGR will be observed, even if half of the survey unit concentration distribution is above the background median by more than the $DCGL_W$. This likelihood is higher when the measurement standard deviation is

large compared to the width of the gray region.

If, as in Scenario A, we assume that the data are approximately normally distributed, the width of the gray region, $\Delta/\sigma = (DCGL_W - LBGR)/\sigma$, can be used to estimate the parameter P_r:

$$P_r = \text{Probability}(U = X - Y > LBGR)$$

$$= \int_{LBGR}^{\infty} \left[\int_{-\infty}^{\infty} f_X(u + y) f_Y(y) dy \right] du$$

$$= \int_{LBGR}^{-\infty} \left[\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-(u+y-DCGL_W-BKGD)^2/2\sigma^2} \frac{1}{\sqrt{2\pi}\sigma} e^{-(y-BKGD)^2/2\sigma^2} dy \right] du$$

$$= \frac{1}{\sqrt{2\pi}\sqrt{2}\sigma} \int_{LBGR}^{\infty} e^{-(u-DCGL_W)^2/4\sigma^2} du$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{DCGL_W - LBGR}{\sqrt{2}\sigma}} e^{-x^2/2} dx$$

$$= \Phi\left(\frac{\Delta}{\sqrt{2}\sigma} \right) \tag{9-9}$$

This is the same as Equation 9-7. Even though the definition of P_r has changed, its value as a function of Δ/σ has not changed. Therefore, the value of P_r computed from Equation 9-7, as found in Table 9.4, can be used in Scenario B as well as in Scenario A. Figure 9.6, expressing the dependence of sample size on Δ/σ is unchanged, and that is why only one version of Table 3.3 is needed for both scenarios.

Values of P_r for distributions other than normal can be calculated from the following equation:

$$P_r = \text{Probability}(U = X - Y > LBGR) = \int_{LBGR}^{\infty} \left[\int_{-\infty}^{\infty} f_X(u + y) f_Y(y) dy \right] du \tag{9-10}$$

where Y is a random measurement from the referenece area with density f_Y and X is a random measurement from the survey unit with density f_X.

As in Scenario A, it may be possible to estimate P_r directly from remediation control survey data. To do this, the data should first be shifted so that the median is at the $DCGL_W$ before the proportion that fall above the LBGR is calculated.

P_r may also be estimated from the odds that a random measurement will be greater than the LBGR versus that it is below the LBGR *when the median is near the $DCGL_W$.* If these odds are $r_1:r_2$, then $P_r = r_1/(r_1+r_2)$. For example, if the odds that a random measurement is greater than the LBGR when the median is near the $DCGL_W$, are 3:1, then $P_r = 3/(1+3) = 3/4 = 0.75$. Notice that since we are assuming that the true median is near the $DCGL_W$, the odds must be greater than 1:1. However, once a survey unit has been remediated, it may be somewhat unnatural to try to estimate the odds this way.